Effectiveness of Community-Based Treatment for Substance-Abusing Adolescents: 12-Month Outcomes of Youths Entering Phoenix Academy or Alternative Probation Dispositions

Andrew R. Morral, Daniel F. McCaffrey, and Greg Ridgeway
RAND

Whereas strong efficacy research has been conducted on novel treatment approaches for adolescent substance abusers, little is known about the effectiveness of the substance abuse treatment approaches most commonly available to youths, their families, and referring agencies. This report compares the 12-month outcomes of adolescent probationers ($N = 449$) who received either Phoenix Academy, a therapeutic community for adolescents that uses a treatment model that is widely implemented across the U.S., or an alternative probation disposition. Across many pretreatment risk factors for relapse and recidivism, groups were well matched after case-mix adjustment. Repeated measures analyses of substance use, psychological functioning, and crime outcomes collected 3, 6, and 12 months after the baseline interview demonstrated that Phoenix Academy treatment is associated with superior substance use and psychological functioning outcomes over the period of observation. As one of the most rigorous evaluations of the effectiveness of a traditional community-based adolescent drug treatment program, this study provides evidence that one such program is effective. Implications of this finding for the dissemination of efficacious novel treatment approaches are discussed.

Between 1995 and 1998, the number of substance abuse treatment admissions for youths in the United States rose by 46%, to 138,000 admissions of 12- to 17-year-olds. This growth is almost exclusively attributable to a steady rise in treatment referrals from the criminal justice system (Substance Abuse and Mental Health Services Administration, 2001). Indeed, by 2000, the criminal justice system referred 50% of all adolescent treatment admissions and 55% of all adolescent admissions to long-term residential treatment programs (U.S. Department of Health and Human Services [DHHS], Substance Abuse and Mental Health Services Administration, Office of Applied Studies, 2003). Despite the criminal justice system's growing reliance on substance abuse treatments, few rigorous studies have evaluated the effectiveness of these services. In this article we describe RAND's Adolescent Outcomes Project, which examined the effectiveness of one such program, the Phoenix Academy of Los Angeles.

There is a small but growing literature on adolescent drug treatment interventions that demonstrates the efficacy of, for instance, family therapy, cognitive therapy, behavior therapy, and other interventions (Williams, Chang, & Addiction Centre Research Group, 2000; Winters, 1999). The majority of these studies, however, concern interventions that Weisz, Weiss, and Donenberg (1992) referred to as *research therapies;* that is, the interventions are typically theory driven; manualized; resource intensive; and implemented in research settings characterized by intensive training, supervision, and fidelity monitoring. Although several of these new treatment methods appear to be efficacious in the treatment of adolescents (e.g., Azrin, Donohue, Besalel, Kogan, & Acierno, 1994; Friedman, 1989; Henggeler, Melton, & Smith, 1992), few programs around the country have implemented them. Instead, as noted in a recent Institute of Medicine report, the treatment approaches most commonly available in the United States use approaches that primarily draw on self-help principles derived from recovery communities and the experiential knowledge gained by counselors, many of whom have personal experience with recovery from drug and alcohol dependence (Lamb, Greenlick, & Mc-Carty, 1998). Following the convention established by this Institute of Medicine report, we refer to these widely available residential or outpatient treatment approaches as *community-based treatments,* although many are also provided in institutional settings, such as prisons.

Two community-based adolescent treatment approaches are common in the United States: (a) Minnesota model treatment, an outpatient or residential approach combining a view of substance dependence and recovery steps derived from Alcoholics Anonymous with techniques drawn from individual and group psychotherapy (Winters, Latimer, & Stinchfield, 1999), and (b) therapeutic community treatment, a residential treatment emphasizing

mutual self-help, behavioral consequences, and a shared set of values concerning "right living" (Jainchill, 1997). Whereas in the Minnesota model dependence is viewed as a disease and the counselor as the agent directing the treatment, the therapeutic community approach views dependence as a symptom of more general behavioral and personality problems and the community itself as the key agent of change (De Leon, 1999).

The effectiveness of community-based treatments for adolescents is typically examined in large observational studies that compare drug use and other problem behaviors in the period preceding treatment entry to some point after discharge. These studies, which include the Drug Abuse Reporting Program (Sells & Simpson, 1979), the Treatment Outcomes Prospective Study (Hubbard, Cavanaugh, Craddock, & Rachal, 1985), the National Treatment Improvement Evaluation Study (Gerstein & Johnson, 1999), and the Drug Abuse Treatment Outcomes Study—Adolescent (Hser et al., 2001), tend to confirm that substance use and other problem behaviors diminish after treatment entry.

Because the research designs of observational studies are weak, changes in problem behaviors cannot be attributed to the receipt of treatment, as opposed to maturation of the cohorts, "natural recovery," or regression toward the mean (Shadish, Cook, & Campbell, 2002). To some extent, these confounds could be redressed by comparing the relative effectiveness of different programs. For instance, assuming that youth entering outpatient programs and residential programs are similar and subject to the same forces of maturation, natural recovery, and regression toward the mean, any differences in outcomes associated with outpatient treatment or residential treatment might be considered a candidate treatment effect. Unfortunately, the assumption that youth entering different programs are similar is contradicted by the available data. Each major observational study of adolescent drug treatments has found significant differences among treatment groups on pretreatment characteristics.

Many pretreatment characteristics, such as problem severity, treatment motivation, criminal history, school problems, and social environment factors, reliably predict poor treatment outcomes (Catalano, Hawkins, Wells, & Miller, 1991; Galaif, Hser, Grella, & Joshi, 2001; Latimer, Newcomb, Winters, & Stinchfield, 2000; Melnick, De Leon, Hawke, Jainchill, & Kressel, 1997). Thus, pretreatment differences between treated cohorts render differences in treatment outcomes ambiguous; they could result from true differences in the effectiveness of the treatment modalities, or from differences in expected rates of recidivism, relapse, and other psychosocial outcomes for cohorts with different risk profiles.

Few community-based treatments have been studied using rigorous evaluation designs that control for pretreatment differences between treated and comparison youths. In the most recent and thorough review of adolescent substance abuse treatment outcomes, Williams et al. (2000) identified just 13 studies that evaluated treatment outcomes using a comparison group that received a different treatment or no treatment. Of these, 4 examined a community-based treatment, and just 1 used random assignment. In this study, Amini, Zilberg, Burke, and Salasnek (1982) found no differences in drug use outcomes 1 year after assignment to either a residential treatment program with a psychoanalytic approach or routine outpatient probation supervision and services. However, these outcome analyses included just 73 participants, giving the study low statistical power to detect genuine treatment effects of the magnitude that might be expected.

Three remaining studies of community-based treatments identified by Williams et al. (2000) each used a quasi-experimental design that included a comparison group more or less well matched with the treatment group. Braukmann et al. (1985) constructed two comparison groups for their study of probationers assigned to "teaching family" group homes. The first consisted of undetained friends of the treated youth, and the second consisted of youth assigned to nonteaching family group homes. This design succeeded in constructing comparison groups with pretreatment drug use behaviors comparable to those of the treatment group. Outcomes assessed 3 months after treatment completion revealed no group differences. Again, however, sample sizes for the comparison of the treatment group with the no-treatment comparison group ($n = 16$ for each condition) was so small that this study had little chance of detecting treatment effects even if they were present in expectable magnitudes. Two other studies of community adolescent drug treatment services—Grenier's (1985) evaluation of the Baton Rouge General Hospital's Adolescent Chemical Dependency Unit and Vaglum and Fossheim's (1980) evaluation of an inpatient ward at Dikemark Hospital in Oslo, Norway, serving drug abusers between the ages of 15 and 27—also suffered from methodological problems that limit any conclusions about treatment effectiveness (Catalano et al., 1991; Williams et al., 2000).

More recently, Winters and Latimer and their colleagues have reported a pair of analyses examining the outcomes of youths receiving residential or outpatient Minnesota model treatment at a large community-based program (Latimer et al., 2000; Winters, Stinchfield, Opland, Weller, & Latimer, 2000). In the first analysis, drug use frequency at a 12-month follow-up assessment was examined among youths receiving residential care, outpatient care, or no treatment (a wait list control group). Although random assignment was not used, the authors reported finding no significant differences between the residential and outpatient samples on a range of pretreatment risk factors, including drug use severity, prior treatments, and demographic characteristics. On a smaller number of pretreatment characteristics, Winters et al. (2000) reported no significant differences between the treatment and control groups. Twelve-month posttreatment outcomes for these groups revealed no significant differences between the residential and outpatient cohorts. Youth who received either form of treatment were, however, significantly more likely than untreated control participants to report abstinence at the 12-month follow-up assessment.

In a second analysis from this same study, Latimer et al. (2000) reported a structural equation model fit to data from 225 youths before treatment and 6 and 12 months later. Their model included treatment modality (residential vs. nonresidential), treatment length, gender, and risk and protective factor indices as predictors of 12-month substance abuse problem severity. The selected model revealed no significant effect of treatment modality on 12-month drug problem severity. This study represents an advance in the analysis of the relative effectiveness of different program types, but it is limited because it controlled for pretreatment characteristics using relatively few aggregate indices (the risk and protective factor indices) and just one demographic characteristic (gender).

RAND's Adolescent Outcomes Project examined the effectiveness of a widely available adolescent therapeutic community treatment approach, the Phoenix Academy, as implemented in Los Angeles. The Phoenix Academy treatment model was developed by Phoenix House, a large substance abuse treatment provider, and

has been implemented in 11 programs in seven states. Phoenix Academy of Los Angeles is a 150-bed, residential therapeutic community for adolescents with an on-site school staffed by the Los Angeles County Board of Education (Jainchill, 1997; Morral, Jaycox, Smith, Becker, & Ebener, 2003). During what is planned as a 9- to 12-month treatment, residents progress through phases associated with increasing program privileges (e.g., leave on day passes, possession of personal belongings) and responsibilities (e.g., residents assume increasingly interesting and more responsible job functions within the community). Days are highly structured, with most waking hours spent in school, community meetings, lectures, encounter groups, family or individual counseling, recreation, and other activities. All program elements are guided by a core set of beliefs about addiction, recovery, and "right living" common to most therapeutic community treatments. For instance, the Phoenix Academy model emphasizes honesty, personal responsibility, community involvement, and mutual self-help as key components of the treatment method (see De Leon, 1999; Jaycox, Marshall, & Morral, 2002). Professional program staff include psychiatrists, psychologists, social workers, and counselors. Many staff members are themselves in recovery. A more detailed description of the Phoenix Academy program philosophy, structure, and staffing is available in Jaycox et al. (2002).

In this article, we compare the 12-month substance abuse, psychological functioning, and criminal activity outcomes of 449 adolescent probationers who received treatment at the Phoenix Academy of Los Angeles or some other probation disposition. We used a powerful case-mix adjustment strategy to correct these outcome analyses for pretreatment differences between youths entering the Phoenix Academy and those receiving other probation dispositions, in order to address the question "Do youths who enter Phoenix Academy have better outcomes than the average expected outcome for comparable youth receiving alternative probation disposition?" Because many available alternatives did not include intensive substance abuse treatment, we hypothesized that Phoenix Academy would produce superior drug use, crime, and psychological outcomes.

Method

Participants

Los Angeles has the largest juvenile probation system in the United States. Recruitment for this study occurred in all three Los Angeles juvenile halls, using procedures approved by the juvenile court, probation, and RAND's Human Subjects Protection Committee. All participants were legally wards of the Los Angeles Superior Courts, which provided consent to interview youths in its care who met the study eligibility requirements and offered their own voluntary informed assent. Although the court provided research participation consent, parents of youth were also provided an opportunity to remove their children from the study using a passive informed consent procedure. For 11 months, beginning in February 1999, all detainees meeting eligibility requirements were invited to participate in the study. To increase the Phoenix Academy sample size, recruitment of juveniles assigned to Phoenix Academy was extended for 4 additional months.

To identify comparison individuals likely to have pretreatment risk characteristics similar to those of youths entering Phoenix Academy, we conducted key informant interviews with each of the Probation Zero Incarceration Program (ZIP) officers responsible for making referrals to community placements such as Phoenix Academy. We asked the officers to indicate where they would refer a youth if he or she seemed best suited

for Phoenix Academy but no bed was available there. Although no explicit rules guide these placement decisions, the ZIP officers agreed that there were seven programs to which they were likely to send youths with behavioral profiles like those they sent to Phoenix Academy. Six of these seven programs agreed to participate in this study. These programs proved to be comparable to Phoenix Academy on a range of factors, including size, planned duration, staffing, and probation referral patterns. Although all programs offered some type of substance abuse treatment services, only Phoenix Academy specialized in substance abuse treatment services. A more detailed discussion of program characteristics is available in Morral et al. (2003).

The study population was drawn from all cases referred by probation to any of these seven group homes (Phoenix Academy plus the six others identified by ZIP officers) during the recruitment period. Youths eligible for the study were required to: (a) be between 13 and 17 years old at study entry, (b) provide a written informed assent to participate in the research, and (c) provide permission to notify a parent or legal guardian of study participation. The study excluded youths if: (a) their facility with English was too weak to participate in the English language interviews, (b) they were admitted to a residential program before they could be interviewed by RAND field staff, or (c) a parent requested that his or her child be excluded.

Figure 1 illustrates the flow of participants from potential eligibility through inclusion in the study. A total of 574 youth were potentially eligible to participate because they were referred to one of the seven participating group homes and met the age inclusion criterion. Of these, 125 were not recruited, because they were placed or moved to a different detention facility before field interviewers could meet with them (84%), because they did not speak English (5%), because they refused (2%), or for other miscellaneous reasons (7%). The remaining 449 youths were enrolled in the study as Phoenix Academy condition youths if they were admitted to that program, or as comparison condition youths if they were not. Because not all youths were ultimately admitted to the program to which they were referred, 29.6% of the comparison condition youths entered a total of 22 other residential group home programs; 3.6% received a probation camp disposition; and 8.4% had other dispositions, including home on probation, hospitalization, jail, and absconded before placement. As such, the comparison condition is representative of the universe of probation dispositions experienced by youths meeting the eligibility requirements who do not enter Phoenix Academy.

Study follow-up retention was excellent. At each of the 3-, 6-, and 12-month assessments, more than 90% of the baseline sample ($N = 449$) was located and successfully interviewed (3 months, $N = 406$; 6 months, $N = 410$; 12 months, $N = 408$).

Procedure

RAND interviewers reviewed juvenile hall detention logs daily to identify eligible candidates, whom they approached with details about the study. On receiving informed assent from the youths, the first of four face-to-face interviews occurred immediately in an attorney interview room within the detention facility. Participants were promised confidentiality, and their participation was remunerated with a gift worth $15. Follow-up interviews 3, 6, and 12 months later occurred in locations convenient for the participant that afforded auditory privacy and safety for the interviewer, such as away from others in group home cafeterias, public spaces, interview rooms in jails and detention centers, and restaurants.

Measures

Instrument. The principal data collection instrument at each of the four assessments was a version of the Global Appraisal of Individual Needs (GAIN; Dennis, 1998). The GAIN has established norms for both adults and adolescents (Dennis et al., 2000; Dennis et al., 1999) and contains eight main sections (background, substance use, physical health, risk behaviors, mental health, environment, legal, and vocational). It provides more than
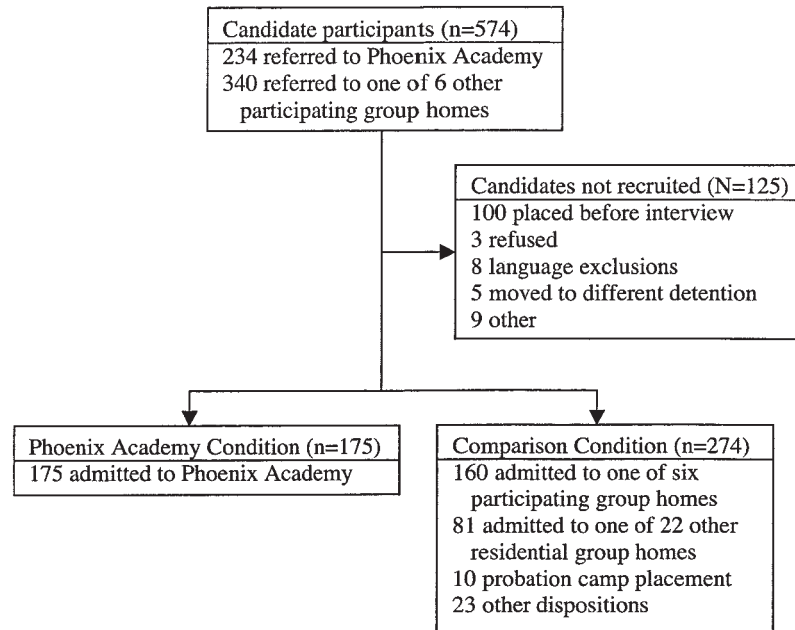
```
┌─────────────────────────────────────────┐
│ Candidate participants (n=574)          │
├─────────────────────────────────────────┤
│ 234 referred to Phoenix Academy         │
│ 340 referred to one of 6 other          │
│    participating group homes            │
└─────────────────────────────────────────┘
```

```
┌──────────────────────────────────────────┐
│ Candidates not recruited (N=125)          │
├──────────────────────────────────────────┤
│ 100 placed before interview               │
│ 3 refused                                 │
│ 8 language exclusions                     │
│ 5 moved to different detention            │
│ 9 other                                   │
└──────────────────────────────────────────┘
```

```
┌─────────────────────────────────────┐    ┌──────────────────────────────────────┐
│ Phoenix Academy Condition (n=175)   │    │ Comparison Condition (n=274)         │
├─────────────────────────────────────┤    ├──────────────────────────────────────┤
│ 175 admitted to Phoenix Academy     │    │ 160 admitted to one of six           │
│                                     │    │    participating group homes         │
│                                     │    │ 81 admitted to one of 22 other       │
│                                     │    │    residential group homes           │
│                                     │    │ 10 probation camp placement          │
│                                     │    │ 23 other dispositions                │
└─────────────────────────────────────┘    └──────────────────────────────────────┘
```

Figure 1. Participant selection and recruitment flowchart.

100 symptom, change score, and utilization indices that have good internal reliabilities (Dennis et al., 2002). The baseline instrument required, on average, 91 min to complete ($SD = 22$). A follow-up version of the GAIN was used at the 3-, 6-, and 12-month interviews. Average completion times for these three follow-up instruments were 64 min ($SD = 32$), 60 min ($SD = 20$), and 54 min ($SD = 20$), respectively.

Outcome measures. Substance use, crime, and psychological distress outcomes were assessed using GAIN scales. The Substance Problem Index is a 16-item symptom count of substance abuse and dependence symptoms listed in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–IV;* American Psychiatric Association, 1994). In the study sample, this scale had a high internal reliability (Cronbach's $\alpha = .92$), and in a test–retest study spanning 90 days, reliability of adolescent reports on this scale was found to be good ($r = .73$; Dennis et al., 2000). The Substance Use Density Scale sums the number of days of reported use of each of 12 classes of drugs (alcohol, marijuana, crack, inhalants, heroin, etc.). Like a "sources of income" scale, there is no expectation that a common factor contributes to the variance in each item, so interitem reliability is neither expected nor assessed. The Substance Involvement Scale assesses the range of substances recently used. Respondents receive 1 point for each of 12 classes of drugs used within the past 90 days. The final substance use outcome is tobacco smoking recency, a single item on which responses are provided on a scale ranging from 0 (*never*) to 6 (*past 2 days*).

Crime outcomes included survey items inquiring about the number of arrests experienced in the past 90 days, the number of days during the past 90 during which the respondent engaged in illegal activities, and the amount of time since the most recent illegal activity committed by the respondent. Three additional outcomes measure the frequency of property, violent, and drug-related crimes in the past 90 days. Each measure sums across multiple specific crimes the number of acts committed; for example, property crime frequency during the past 90 days is the sum of self-reported acts of vandalism, forgery, petty theft, grand larceny, and breaking and entering.

Psychological functioning outcomes were assessed with one item on the recency of psychological distress and three scales based on the Hopkins Symptom Checklist (Derogatis, Lipman, Rickels, Uhlenhut, & Cori, 1974). The Somatic Symptoms Index is a 4-item symptom count assessing so-matic symptoms commonly associated with psychological distress (e.g., headaches, dizziness, sleep problems, dry mouth, and diarrhea). Interitem reliability on this scale was good, especially considering the small number of items on which it is based (Cronbach's $\alpha = .68$). The Depressive Symptoms Index counts the presence of 6 sets of symptoms associated with depression (e.g., feeling sad or depressed, loss of energy, concentration problems, and irritability). Interitem reliability was good (Cronbach's $\alpha = .74$). Finally, the Anxiety Symptoms Index counts the presence of 10 sets of symptoms related to various anxiety disorders (e.g., anxiety, compulsive thoughts and behaviors, restlessness, phobias, etc.). Interitem reliability was again good (Cronbach's $\alpha = .79$).

Pretreatment characteristics. Group differences between Phoenix Academy (PA) and comparison (Comp) condition youths were assessed on a range of pretreatment characteristics, including gender, age, race, lifetime arrests, past-year health, last grade completed, recency of paid work, self-reported treatment need, age at first drug use (or alcohol use to intoxication, whichever came first), number of prior drug treatments, recency of injection drug use, days of alcohol/drug use in the past 90 days, and days drunk or high in the past 90 days. In addition, lifetime versions of the Substance Problem Index and Substance Involvement scales, described above, were included as indicators of lifetime substance use history. Presumptive lifetime substance use disorder diagnoses are calculated by applying algorithms like those in the *DSM–IV* to lifetime symptoms of substance use disorders. These subscales of the Substance Problem Index produce interitem reliabilities with Cronbach's alpha scores above .70 (Dennis et al., 2000).

Statistical Approach

Missing data. No item included in the analyses reported here had more than 4.01% missing data, and on average items had just 1.17% missing data ($SD = 0.73\%$). Despite this low rate, we used a regression model hot-deck imputation procedure (Little & Rubin, 1987) so that scale scores could be based on all scale items. Specifically, for a given variable with missing values, we modeled the expected value of the variable as a function of demographic characteristics, drug use history, psychological status, treatment history, legal history, responses to the same item at earlier survey waves (when available) and other variables from the GAIN (53 or more

items in all) and used the observed data to fit the model. Records were stratified into 10 sets by the percentiles of the predicted scores from this fitted model. For each observation with a missing value, a "donor" value was drawn at random from the records in the same predicted score stratum as the case with the missing value. Data were not imputed for whole surveys that were missed.

Case-mix adjustment. The *propensity score* is a participant's probability of being a member of the PA condition, given a set of observed characteristics or covariates. Rosenbaum and Rubin (1983) showed that, under certain assumptions, conditioning analyses of the propensity score can remove the confound between treatment effects and pretreatment risk factors when comparing groups in observational studies. Analyses may be conditioned on propensity scores by stratifying cases on the propensity score (Rosenbaum & Rubin, 1985) or by weighting cases with the odds associated with their propensity score (Hirano, Imbens, & Ridder, 2003). In the present analyses we adopted the latter approach. Therefore, each Comp case receives an associated weight of $p_i / (1 - p_i)$, where p_i is the propensity score describing the probability that case i belongs to the PA condition. This case weighting approach properly downweights Comp cases with covariates very dissimilar to those of PA cases and upweights Comp cases with covariates similar to PA cases. The weighted Comp cases have covariate distributions that tend to match those of the PA cases.

Using a nonparametric logistic regression procedure, we estimated the probability of assignment to the treatment group from 41 pretreatment characteristics selected *a priori* to represent a broad set of pretreatment risk factors. These included baseline values on all outcome measures described above as well as demographic characteristics (gender, age, and race), school and work participation, current drug use and drug problems (e.g., withdrawal symptoms, self-reported treatment need), drug use history (e.g., age at first use, prior drug treatments, injection drug use recency), criminal history (e.g., lifetime arrests, days in a controlled environment), treatment readiness (e.g., measures of treatment motivation and treatment resistance), measures of the social environment from which youths have arrived in detention, and measures of physical and mental health (e.g., ratings of past-year health, symptoms of attention-deficit/hyperactivity disorder).

The resulting propensity score model explained 41% of the variance in assignment, with six variables accounting for more than 50% of explained variance: (a) a treatment motivation index; (b) self-reported need for treatment; (c) the Substance Use Density Index; and an index of recent behavioral problems associated with (d) attentional, (c) hyperactivity, and (e) conduct disorders. A detailed report on the propensity score regression modeling approach we used in these analysis is available in McCaffrey, Ridgeway, and Morral (in press).

Before applying weights to the Comp condition, we observed significant group differences on many pretreatment characteristics. For instance, 13 of the 16 pretreatment characteristics listed in Table 1 were significantly different using an alpha of .05, and 7 of the 14 outcome variables assessed at baseline were significantly different. In general, PA youths were older; had completed more school; were more likely to be White and female; and were more involved with drugs, alcohol, and drug-related crimes. Comp youth were more likely to report better past-year health, fewer prior drug treatments, and an earlier age at first drug use. Not surprisingly, PA youth were more likely to report needing treatment for marijuana and other drug use.

After we applied weights derived from the propensity scores to Comp cases, between-group differences diminished substantially. For instance, the number of significant differences between the PA condition and the weighted Comp condition reduced to just 3 the 16 background characteristics listed in Table 1, and no significant differences remained among the baseline values of the 14 outcome variables. More important, the average magnitude of the t scores described in Table 1 drops substantially after weighting, from 3.24 to just 1.09. The weighting failed to eliminate differences for gender, last grade completed in school, and injection drug use recency. Although this number of significant differences is small, and well within the number expected by chance when using an alpha of .05

across 30 t tests, we include these 3 variables as covariates in each of the outcomes analyses described below.

Comparison of t scores before and after weighting is somewhat misleading. The variance of the weighted mean for the Comp group is greater than the variance for the unweighted mean (the increase is roughly a factor of 2), so comparisons after weighting have less power to detect significant differences. To better establish the magnitude of the effect produced by the propensity score case weights, McCaffrey et al. (in press) examined pseudo t tests of the group differences before and after weighting. These tests hold the denominator of the t statistic constant before and after weighting at the (unweighted) PA standard error. Changes in this pseudo t statistic therefore reflect only the magnitude of the change in the Comp group means on each variable. Across the 36 comparisons in Table 1, Comp group case weights have the effect of reducing pseudo t scores from an average magnitude of 4.08 to just 1.78. Thus, the case weights have the effect of making the Comp condition substantially more like the PA condition on a wide range of pretreatment characteristics.

Treatment outcome analyses. We assessed PA and Comp condition outcomes using a repeated measures analysis that included one between-group factor (condition) and four levels of time (baseline, 3 months, 6 months, and 12 months). As recommended by Shadish et al. (2002), we included as covariates in these analyses the three pretreatment characteristics that significantly differentiated conditions after weighting: (a) last grade completed, (b) injection drug use recency, and (c) gender. All 90-day frequency variables, symptom count, and crime count indices with a range greater than 15 received square-root transformations to correct distributional skew.

On the basis of preliminary examination of the outcomes, we modeled outcome time trends for each treatment group as a piecewise linear function with one linear trend from baseline to 3 months and a second contiguous linear trend from 3 to 12 months. There was no significant lack of fit in this model compared to a model with separate trends between each two time points. The analyses were implemented in SAS PROC MIXED (SAS Institute Inc., 2001).

Analyses of relative treatment effects were conducted in two stages. First, we performed multivariate repeated measures analyses on three families of outcome variables (substance use, crime, and psychological functioning). For these analyses, significant three-way interactions of outcome family × treatment condition × the piecewise linear time trends indicate that the piecewise linear time trends for at least one outcome variable differ between the two treatment conditions. To identify the variables and time points producing significant multivariate interactions, in the second stage of the analysis we conducted univariate repeated measures analyses in which we examined all significant Condition × Time interactions for evidence of differential outcomes. We used the two-stage approach to reduce the probability of Type I errors within groups of outcomes by requiring that the group of interactions be significant before we explored each individual outcome (Steel & Torrie, 1980). In addition, we interpreted significance levels using Hochberg's (1988) step-down Bonferroni correction for Type I error.

Results

Unless otherwise noted, all results reported in this section are from analyses with the weighted comparison group to reduce confounding of baseline risk factors and program effects.

Participants

Several demographic, drug use, crime, arrest, education and other pretreatment characteristics of the weighted and unweighted treatment conditions are presented in Table 1. In addition, 43% of the PA and 35% of the Comp condition participants lived with a parent in the year prior to study entry, $T(447) = 1.39$, *ns.* Mem-

Table 1

Comparison of Baseline Characteristics: Phoenix Academy (PA) and Comparison Condition (Comp) Before and After Weighting the Comparison Condition (Comp-WT)

	PA		Comp		Comp-WT		PA vs. Comp	
							Unweighted	Weighted
Baseline characteristics	M	SE	M	SE	M	SE	T(447)	T(447)
Demographics								
Female (%)	18.29	2.93	9.12	1.74	7.30	1.90	−2.86**	3.17**
Age (years)	15.82	0.07	15.31	0.08	15.71	0.09	−4.57***	1.04
Race (%)								
African American	8.57	2.12	18.61	2.36	12.20	2.10	2.95**	−1.19
Latino/Hispanic	60.00	3.71	52.19	3.02	54.20	4.40	−1.62	1.00
White	20.57	3.06	13.14	2.04	18.00	4.00	−2.10***	0.50
Lifetime arrests[a]	1.82	0.07	1.85	0.09	1.96	0.10	0.25	−1.19
Past-year health[b]	1.87	0.08	1.55	0.07	1.66	0.10	−3.05**	1.69
Last grade completed	9.04	0.09	8.59	0.08	8.76	0.11	−3.51***	1.99*
Recency of paid work[c]	1.32	0.11	1.13	0.08	1.13	0.11	−1.42	1.24
Drug use history								
Substance disorder (%)								
Physical dependence	60.00	3.71	37.23	2.93	58.40	4.10	−4.83***	0.29
Dependence	10.29	2.30	6.20	1.46	7.30	2.20	−1.57	0.93
Abuse	23.43	3.21	27.01	2.69	21.60	3.00	0.85	0.43
Substance Problem Index (lifetime)[a]	3.05	0.06	2.22	0.08	2.88	0.07	−7.59***	1.88
Substance Involvement Scale (lifetime)[a]	2.15	0.05	1.73	0.04	2.03	0.05	−6.41***	1.71
Self-reported treatment need for (%)								
Alcohol	4.57	1.58	5.47	1.38	7.50	2.40	0.42	−1.02
Marijuana	27.43	3.38	12.04	1.97	22.70	4.00	−4.21***	0.90
Other drugs	32.00	3.54	6.93	1.54	20.70	4.70	−7.33***	1.91
Age at first use	12.55	0.14	11.97	0.19	12.25	0.21	−2.22*	1.19
Prior drug treatments[a]	0.98	0.13	0.52	0.07	0.90	0.12	−3.30***	0.45
Injection drug use recency[c]	1.85	0.16	1.22	0.16	1.22	0.16	−3.84***	2.74**
Days of alcohol/drug use (in past 90)[a]	6.19	0.24	3.77	0.22	5.52	0.26	−7.25	1.87
Days drunk/high (in past 90)[a]	4.18	0.26	2.51	0.19	3.91	0.29	−5.28***	0.69
Baseline performance on outcomes								
Drug use								
Substance Problem Index (past 90 days)[a]	1.61	0.10	0.70	0.07	1.35	0.12	−7.91	1.67
Substance Involvement Scale (past month)[a]	0.86	0.06	0.40	0.04	0.71	0.08	−6.45***	1.43
Substance Use Density Index[a]	7.61	0.33	4.59	0.29	6.90	0.40	−6.81***	1.39
Smoking recency[c]	2.93	0.12	2.25	0.10	2.77	0.15	−4.31***	0.83
Psychological								
Psychological distress recency[c]	1.14	0.12	1.30	0.11	1.42	0.15	0.95	−1.48
Somatic Symptoms Index	1.14	0.10	0.92	0.07	1.02	0.09	−1.88***	0.89
Depressive Symptoms Index	2.39	0.14	2.05	0.11	2.36	0.17	−1.87	0.15
Anxiety Symptoms Index	2.82	0.19	2.61	0.15	2.83	0.19	−0.85	−0.01
Crime								
Crime recency[c]	2.54	0.11	2.58	0.09	2.69	0.11	0.29	−0.91
Crime days (in past 90)[a]	4.26	0.26	3.20	0.20	4.25	0.31	−3.28**	0.02
Arrests (in past 90 days)[a]	0.76	0.05	0.70	0.04	0.73	0.05	−1.03**	0.39
Property crimes (in past 90 days)[a]	1.90	0.19	1.65	0.17	1.91	0.27	−0.98	−0.04
Drug crimes (in past 90 days)[a]	1.60	0.22	1.12	0.15	1.59	0.25	−1.85	0.03
Violent crimes (in past 90 days)[a]	0.98	0.11	1.08	0.09	1.12	0.12	0.63***	−0.83

Note. Sample sizes: PA *n* = 175; Comp *n* = 274; effective sample size after weighting, Comp-WT *n* = 127.64. Percentages do not sum to 100% because each variable has one holdout category. Holdout categories are male (gender), other (race), none (substance disorder), none (self-reported treatment need). [a] Past 90-day frequency and count variables with a range greater than 15 were transformed to their square roots to reduce variable skew. [b] Past-year health scale ranged from *excellent* (0) to *poor* (4). [c] Recency scale spans 0 (*never*) to 6 (*past 2 days*).
* *p* < .05. ** *p* < .01. *** *p* < .005.

bership in a street gang was acknowledged by 30% of PA and 34% of Comp participants, *T* = −0.79, *ns.*

Treatment Exposure and Time at Risk

Ideally, we would compare outcomes associated with PA treatment to those for similar youths who receive no treatment. In this study, however, most Comp youth (78.9%) were admitted to another residential group home after their baseline interview. The

remaining cases were sent to a probation camp placement (4.1%) or had various other dispositions (17.0%).

Differences in treatment outcomes may be associated with differences in exposure to treatment rather than differences in the relative effectiveness of each treatment condition. Thus, we assessed treatment exposure for youths in the PA and Comp conditions using self-reports of the number of days spent in residential treatment centers. Neither significant nor systematic group differences were observed on this variable at baseline, Month 3, Month

6, or Month 12. Mean days in a residential care facility across the four assessments were 6.7, 55.2, 49.5, and 34.3, respectively, for the PA youths, compared with 9.3, 50.1, 46.2, and 32.8 for Comp youth, $t(447) = -1.22$, $t(404) = 0.97$, $t(408) = 0.63$, and $t(406) = 0.31$. Thus, youths in each condition appear to have received comparable exposures to residential programs.

It is important to note, however, that the survey item on days in residential care concerns any group home setting, not necessarily the residence to which the individual was first assigned. Nevertheless, the mean length of stay at Phoenix Academy for youths in the PA condition was 161.9 days ($SD = 131.0$), which was quite similar to the mean length of first placement for those youths in the Comp condition who entered a residential treatment program ($M = 169.3$ days, $SD = 132.2$). Similarly, as of the 12-month assessment, the percentage of youths remaining in the group home to which they were first assigned was 19.2% for the PA condition and 21.8% for the Comp condition. Thus, program retention for PA and for Comp youths is similar.

Crime and substance use outcomes may be confounded with time "at risk"; that is, time spent in environments where drug use and crime are most likely to occur. If, for instance, one condition is associated with higher rates of subsequent detention (a poor outcome), this could result in lower rates of crime and substance use (a seemingly good outcome) simply because youths were less able to engage in drug use and crime while in detention. Thus, we also compared conditions on the amounts of time youths reported being in controlled environments. Specifically, at each assessment youths were asked "During the past 90 days how many days did you live someplace . . . where you were not free to come and go as you please—such as a juvenile hall, an inpatient program, or a hospital?" Prior to treatment entry, youths reported similar num-

bers of days in a controlled environment (Ms = 15.8 and 16.7 days for PA and Comp, respectively; $t[447] = -0.33$, ns). Three months after study entry, youths again reported comparable numbers of days in controlled environments (PA = 65.4, Comp = 63.4, $t[404] = 0.43$, ns). By the 6-month assessment, group means diverged some (PA = 61.1, Comp = 48.9, $t[408] = 2.39$, $p < .05$), although no significant differences were found at the 12-month assessment (PA = 37.9, Comp = 43.3, $t[406] = -1.10$, ns). Thus, youths in both conditions reported comparable exposure to substance use and crime risk as measured by time in controlled environments.

Outcome Analyses

The entire set of Condition × Piecewise Linear Time Trend interaction F tests in the case-mix adjusted outcomes model, and the effect sizes associated with the single-degree-of-freedom tests, are displayed in Table 2. Effect sizes are analogous to Cohen's d (Cohen, 1988): The effect size for change from baseline to first follow-up (Time 0–Time 1) equals the model-based estimate for the difference between PA and Comp youths in the change in the means from baseline to first follow-up, divided by the model-based estimate of the pooled within-group standard deviation of baseline-to-first-follow-up change scores. Similarly, effect sizes for Time 1 to Time 3 and for Time 0 to Time 3 are based on model estimates of the mean and standard deviation of the corresponding difference scores.

The case-mix adjusted outcome trends, which may be used to interpret the direction of these interactions, are displayed in Figure 2. To enhance comparability of the figures' scales, each displays an outcome variable plotted against an ordinate spanning

Table 2
Condition × Time Interaction Significance Tests (F Tests) for Multivariate and Univariate Hypotheses and Effect Sizes (Cohen's d)

| | | | Single-*df* tests | | | | | |
| | | | Time 0 to Time 1 | | Time 1 to Time 3 | | Time 0 to Time 3 | |
Variable	Multivariate F[a]	Univariate F[b]	d	F	d	F	d	F
Drug use	3.84***							
Substance Problem Index (past month)		3.02*	−.20	3.28	−.10	0.72	−.27	5.76*
Substance Use Density Index		4.10*	−.32	5.92*	.04	0.07	−.25	5.29*
Substance Involvement Scale (past 90 days)		3.42*	−.29	5.47*	.02	0.02	−.24	4.285*
Tobacco smoking recency		6.80***	−.30	6.15*	.45	12.43***	.20	2.822
Psychological characteristics	2.09*							
Somatic Symptoms Index		4.12*	−.06	0.26	−.27	4.30*	−.32	7.344**
Depressive Symptoms Index		4.17*	.07	0.30	−.34	8.01***	−.22	3.168
Anxiety Symptoms Index		5.52***	.04	0.12	−.39	10.34***	−.29	6.101*
Recency of psychological distress		0.66	−.05	0.16	−.11	0.80	−.15	1.124
Crime	0.74							
Arrests (past 90 days)		0.06	−.04	0.12	.01	0.01	−.03	0.053
Property crimes (past 90 days)		0.42	.02	0.02	−.12	0.79	−.08	0.449
Violent crimes (past 90 days)		1.73	−.15	1.79	−.10	0.55	−.21	3.098
Drug crimes (past 90 days)		0.06	−.04	0.09	.05	0.10	.00	0
Crimes days (past 90 days)		0.76	−.01	0.01	−.16	1.12	−.13	1.21
Recency of illegal behavior		1.34	.07	0.38	−.20	2.56	−.15	1.932

Note. All analyses included as covariates the three pretreatment characteristics found to significantly differentiate pretreatment groups after weighting: gender, injection recency, and highest grade completed. [a] Numerator $df = 8$ for the substance and psychological classes; $df = 12$ for criminal class. Denominator $df = 448$ for all classes. [b] $df = 2$ and 444 for all tests.
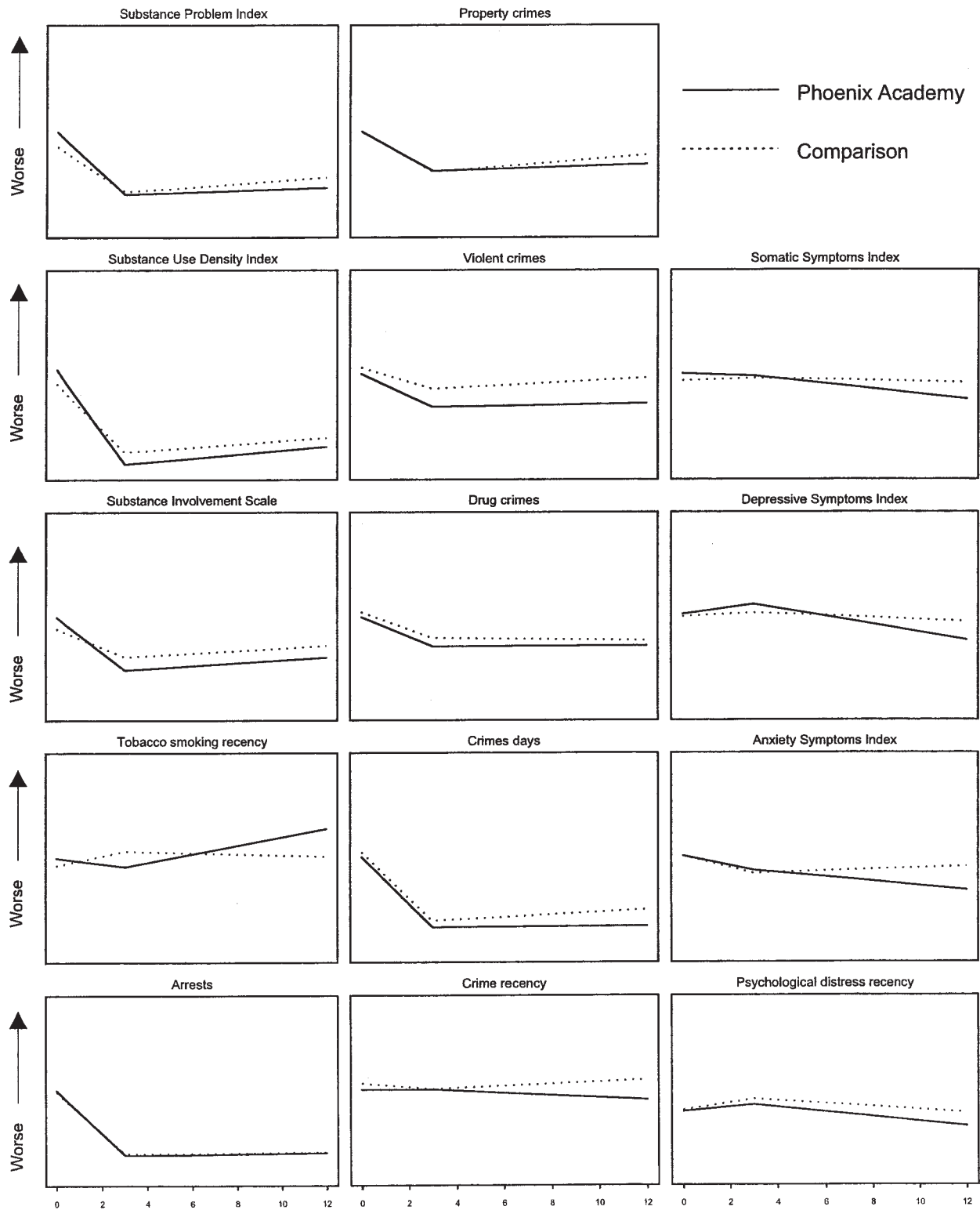* $p < .05$. ** $p < .01$. *** $p < .005$.

Figure 2. Case-mix adjusted time trends for Phoenix Academy and the weighted Comparison condition on each of the 14 outcome variables.

±1.28 standard deviation units around the PA baseline. Thus, each figure's range is from the 10th to the 90th percentile of the baseline PA responses if they are normally distributed. The abscissa depicts time and runs from the baseline interview (at Time 0) to the final, Month 12 interview. In each of these standardized figures, lower values represent more desirable outcomes.

Substance Use Outcomes

PA youth had significantly different substance use trajectories in the year following admission than Comp youths. The multivariate test of the Condition × Time interaction for the set of substance use outcomes was significant, as were univariate tests for each of the substance use scales. Figure 2 and the single-degree-of-freedom tests in Table 2 illustrate that the Substance Use Density Index, the Substance Involvement Scale, and smoking recency all showed significantly greater problem reductions in the first 3 months for PA youths than for Comp youths; moreover, all but one of these problem areas remained lower for PA youths than for Comp youths in the ensuing 9 months. Thus, for the Substance Problem Index, the Substance Use Density Index, and the Substance Involvement Scale, PA youths showed the same trend toward increasing use in Months 3 to 12 as did Comp youths; however, because PA problems did not accelerate faster than Comp, 12 months after intake the PA youth outcomes were significantly better than Comp youth outcomes on all three substance use measures, the size of these effects all being around −.25, conventionally referred to as between a small and medium sized effect (Cohen, 1988).

Smoking recency worsened for PA youths relative to Comp youths during the Month 3 to Month 12 period, a trend that more than eliminates the relative advantages of PA on this outcome during the first 3 months after the baseline interview. These contradictory effects result in a small PA effect on change in smoking recency from baseline through the 12-month follow-up assessment that fell short of statistical significance ($p = .09$).

Psychological Functioning Outcomes

PA was associated with significant psychological benefits over time relative to Comp (see Table 2 and Figure 2). Univariate tests of the Condition × Time interaction reveal that trends differed significantly on three of the four psychological functioning outcomes. For each of these outcomes, significant univariate differences of between small and medium effect sizes appear in symptoms experienced between the 3- and 12-month assessments. Thus, although no significant group differences were found in the first 3 months after treatment entry, PA youths reported significantly greater reductions in psychological symptoms during Months 3 through 12. These differences resulted in significant baseline-to-12-month treatment effect sizes favoring PA for the somatic and anxiety symptom indices, which are small to medium in magnitude. Although the PA effect on the depressive symptom index was also between small and medium size, the *F* test fell short of significance ($p = .07$). Recency of psychological distress did not differ significantly by treatment condition, although the pattern of results was similar to those observed in the symptom indices. Specifically, PA youths had a greater mean reduction in psychological distress recency than Comp youths.

Crime Outcomes

Statistical tests of the Treatment × Time interaction revealed no differences in crime outcomes at the multivariate or univariate levels. Although not significantly different, it is interesting to note that all six of the crime measures exhibited similar patterns of divergence in mean group outcomes. Specifically, PA youth showed greater declines in mean scores over the 12 months of observation than Comp youth on arrests, property crimes, violent crimes, crime days (in the past 90 days), and crime recency, and declines that were equivalent to those observed in Comp youths for drug crimes. In the case of violent crimes, the effect size was small to medium, and the effect approached significance ($p = .08$). This consistent pattern raises the possibility that true treatment effects on crime were present but were too small on our outcome measures to be distinguished from the null hypothesis of no treatment effect.

Effects of Case-Mix Adjustment

To explore the effect of our case-mix adjustment strategy, we reran each of the outcome models but without using case weights to adjust for pretreatment group differences. No significant effect reported in Table 2 failed to attain significance in the unadjusted models. In addition, however, we observed significant Condition × Time effects for the multivariate test of the set of criminal outcomes, for the univariate test of crime days, and for the single-degree-of-freedom test for differences in outcomes on the Substance Problem Index between admission and Month 3. Therefore, our case-mix adjustment had the effect of reducing the apparent differences in outcomes experienced by those who did and did not enter Phoenix Academy.

Discussion

Few prior studies have carefully examined the effectiveness of the types of adolescent treatment services commonly available to youths in the United States. The results of this study suggest that one such program that uses a treatment model widely implemented across the United States, the Phoenix Academy, is associated with better outcomes than the average expected outcome had the same youths received an alternative probation dispositions. As one of the most rigorous evaluations yet reported on the effectiveness of a community-based adolescent drug treatment approach, this study provides evidence that some such programs for adolescents can be effective.

The PA and Comp conditions were well matched in this study after case-mix adjustment. At the baseline assessment, mean scores on a wide range of pretreatment risk factors and all outcome variables were remarkably similar between conditions (see, e.g., Time 0 data for each scale in Figure 2). Similarly, youths were at risk (or, conversely, in controlled environments) for comparable lengths of time during the follow-up interval.

Despite the apparent similarities in the recidivism and relapse risk profiles of youth in each treatment condition, a clear pattern of divergence in outcomes emerged during the 12 months of study observation. Specifically, PA youths had significantly better outcomes for most substance use and psychological functioning outcomes, with the PA condition associated with between small and medium sized effects on these outcomes. In contrast, crime out-

comes were not significantly different between conditions. Our failure to detect crime effects admits many possible interpretations, including that the measures of crime we used are insensitive to the true treatment effects on crime, that Phoenix Academy and the comparison condition interventions were equally effective in reducing crime, and that true differences in treatment effects on crime may be undetectable until youths have been at risk in the community for longer periods.

The possibility that our analyses merely lack the statistical power to detect true treatment effects on crime is suggested by the remarkable consistency of all crime outcomes. Specifically, PA youths had reductions in crime problems that were greater than those of Comp youth for every crime outcome but one in which reductions were equivalent, although none of these differences attained statistical significance. This pattern matches that found for all substance use and psychological functioning outcomes (except smoking recency), for most of which significant differences between conditions were detected.

Treatment entry was associated with sharp reductions in substance use frequency, substance use problems, the range of substances used, and crime. This pattern can be explained, in part, by the fact that youths spent a larger proportion of their first 3 months in controlled environments where substance use and crime were, presumably, more difficult to engage in than during the remaining 9 months, when many had absconded or entered less controlled environments or phases of treatment. Indeed, as the proportion of days in controlled environments decreased, substance use and some crime outcomes became increasingly problematic, although they did not return to pretreatment levels for either PA or Comp youths.

Tobacco smoking recency represents an important exception to this pattern of findings. Despite spending nearly two thirds of their first 3 months in controlled environments, only PA youths reported decreases in tobacco smoking recency. Indeed, Comp youths reported smoking more recently at the 3-month assessment than at the baseline assessment, suggesting that the "controlled environments" in which they found themselves were not effectively preventing tobacco use.

Whereas Phoenix Academy initially appears effective at reducing smoking relative to the comparison condition, all such gains are lost during the subsequent 9 months as more youths leave the program. The relative increase in smoking recency among PA youths presents a striking contrast to their relative reductions in other substance use and problems. This finding highlights the need for two types of study. First, tobacco treatment and prevention programming in substance abuse treatment programs may require improvement (Hahn, Warnick, & Plemmons, 1999; Myers, 1999). Second, adolescents may face a drug use recovery environment that facilitates tobacco use (Bobo, Slade, & Hoffman, 1995). Insofar as cigarette smoking is especially dangerous for young people (DHHS, 1994), efforts to reduce the association between drug use recovery and tobacco use are needed (Goldsmith & Knapp, 1993).

Whether youths are in controlled environments probably has less influence on the expression of psychological distress and symptoms. Predictably, therefore, no sharp changes in these outcomes were observed in the first 3 months of observation. Thereafter, however, PA youths reported progressive reductions in symptoms of psychological distress, while the Comp group symptoms remained relatively stable. It is interesting that this apparent

treatment effect seemed to build even after many or most youth were no longer receiving Phoenix Academy treatment. This may suggest that Phoenix Academy treatment fosters coping strategies or helps youths develop other internal resources on which they successfully draw even after they return to the environments that originally contributed to their psychological distress. As psychological functioning and substance use may be causally related constructs, the psychological-distress findings may also reflect the relative drug use reductions noted among PA youths.

As designed, our analyses support the conclusion that for youths who are likely to be admitted to Phoenix Academy following a probation referral, those actually admitted to Phoenix Academy may be expected to have superior drug use and psychological functioning outcomes after 12 months. This does not imply, however, that Phoenix Academy treatment is superior to any particular alternative probation disposition. We cannot assess that issue because we lack sufficient numbers of cases in each alternative disposition to support such analyses. Similarly, it is quite possible that although Phoenix Academy produces better outcomes for youths like those typically seen at Phoenix Academy, alternative dispositions might better meet the needs of their own usual clientele, even though, on average, these alternative dispositions do not produce superior outcomes for the subset of their youths with pretreatment characteristics like those seen at Phoenix Academy.

Several limitations of this study should be noted. Chief among these is the possibility that the Comp youths differed in important and unobserved ways from PA youths. Because we adopted a case-mix adjustment approach rather than random assignment to conditions, we cannot be certain that any observed differences in outcomes are attributable to treatment rather than to systematic differences in youth risk factors that might have predated treatment. Nevertheless, we note that our case-mix adjustment model was unusually successful at reducing or eliminating group differences in risk factors observed at baseline. Whereas before weighting we observed significant group differences on 20 of the 30 items listed in Table 1, after applying case weights only three significant differences remained, roughly the number expected to be significant by chance when using our alpha of .05. More important, the average magnitude of the t scores reported in Table 1 drops almost two thirds, from 3.24 to 1.09, after weighting. A strength of this study is that in addition to performing a powerful case-mix adjustment model, we report the resulting effects of the model on a wide range of pretreatment characteristics, including all baseline values for each outcome variable.

A second limitation is that we compared the outcomes of PA youths not to a cohort of untreated youths, but to youths who in many cases received active treatments. If both Phoenix Academy and the programs serving youths in the comparison condition had substantial and positive treatment effects of roughly equivalent magnitudes, this would register in our model as an observation of no difference in outcomes between conditions. With our design, we can comment not on the absolute treatment effect but only on the apparent effect relative to that of the comparison condition. This sets a difficult standard for demonstrating program effectiveness and likely results in a misleadingly conservative characterization of Phoenix Academy effectiveness. On the other hand, if not placed in the Phoenix Academy, youth like those in our study are typically placed by the juvenile justice system into some alternative residential program. Thus, our comparison is quite

relevant for probation officials trying to determine the best facilities to place youthful offenders.

Several other limitations common to most research in this area also bear mention. Most of the data used in our analyses were collected through the self-reports of delinquent youths. Self-reports are subject to a number of well-known biases (Morral, McCaffrey, & Iguchi, 2000; Sudman, Bradburn, & Schwarz, 1996). For the purposes of the analyses reported in this study, however, biases in self-reports should affect only conclusions about outcome differences to the extent that youth in one condition are more or less biased in their reporting. We know of no reason to suspect that biases vary by condition.

As discussed earlier, community-based substance abuse treatment services have received less rigorous evaluations than have more recent and novel approaches. This may be due in part to concerns about the generalizability of any evaluation of a community-based treatment. Whereas the approaches more commonly subject to rigorous evaluation are provided with ongoing and intensive supervision and training to ensure implementation fidelity (e.g., Azrin et al., 1994; Dennis et al., 2002; Friedman, 1989; Henggeler et al., 1992), traditional community-based programs rarely have manuals or the fidelity monitoring necessary to support claims of the generalizability of their results. Moreover, these programs often present moving targets, with services and staffing changing along with changes in the treatment funding environment. Whether novel manual-guided approaches would be more stable, or more effective, than traditional treatment approaches when they are implemented in community settings remains to be seen. The results of the present study suggest, however, that some traditional approaches appear to be associated with improvements in drug use, psychological, and perhaps crime-related outcomes beyond what would be expected from residential care generally.

These findings suggest that before substituting the treatment as usual of community-based providers with novel efficacious interventions, more study is required to ensure that the novel approaches are, in fact, superior to the effective traditional ones. Such analyses require a thorough evaluation of current practices of adolescent treatment providers and their effectiveness. This could provide the basis for the establishment of an empirically derived set of adolescent substance abuse treatment best practices and would, at the same time, establish the traditional treatment approaches against which the effectiveness of novel treatments should be compared.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Amini, F., Zilberg, N. J., Burke, E. L., & Salasnek, S. (1982). A controlled study of inpatient vs. outpatient treatment of delinquent drug abusing adolescents: One year results. *Comprehensive Psychiatry, 23,* 436–444.

Azrin, N. H., Donohue, B., Besalel, V. A., Kogan, E. S., & Acierno, R. (1994). Youth drug abuse treatment: A controlled outcome study. *Journal of Child and Adolescent Substance Abuse, 3,* 1–16.

Bobo, J. K., Slade, J., & Hoffman, A. L. (1995). Nicotine addiction counseling for chemically dependent patients. *Psychiatric Services, 46,* 945–947.

Braukmann, C. J., Bedlington, M. M., Belden, B. D., Braukmann, P. D., Husted, J. J., Ramp, K. K., et al. (1985). Effects of community-based group-home treatment programs on male juvenile offenders' use and abuse of drugs and alcohol. *American Journal of Drug and Alcohol Abuse, 11,* 249–278.

Catalano, R. F., Hawkins, J. D., Wells, E. A., & Miller, J. (1991). Evaluation of the effectiveness of adolescent drug abuse treatment, assessing the risk for relapse, and promising approaches for relapse prevention. *International Journal of the Addictions, 25,* 1085–1140.

Cohen, J. (1988). *Statistical power analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

De Leon, G. (1999). Therapeutic communities. In P. J. Ott, R. E. Tarter, & R. T. Ammerman (Eds.), *Sourcebook on substance abuse: Etiology, epidemiology, assessment, and treatment* (pp. 121–136). Needham Heights, MA: Allyn & Bacon.

Dennis, M. L. (1998). *Global Appraisal of Individual Needs (GAIN) manual: Administration, scoring and interpretation.* Bloomington, IL: Lighthouse.

Dennis, M. L., Babor, T. F., Diamond, G., Donaldson, J., Godley, S. H., Titus, J. C., et al. (2000). *The Cannabis Youth Treatment (CYT) experiment: Preliminary findings.* Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration, Department of Health and Human Services.

Dennis, M., Scott, C., Godley, M., & Funk, R. (1999). *Comparisons of adolescents and adults by ASAM profile using GAIN data from the Drug Outcome Monitoring Study (DOMS): Preliminary data tables.* Retrieved June 10, 2003, from http://www.chestnut.org/li/posters/asamprof.pdf

Dennis, M. L., Titus, J. C., Diamond, G., Donaldson, J., Godley, S. H., Tims, F. M., et al. (2002). The Cannabis Youth Treatment (CYT) experiment: Rationale, study design, and analysis plans. *Addiction, 97*(Suppl. 1), 84–97.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhut, E. H., & Cori, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report inventory. *Behavioural Science, 19,* 1–15.

Friedman, A. S. (1989). Family therapy vs. parent groups: Effects on adolescent drug abusers. *American Journal of Family Therapy, 17,* 335–347.

Galaif, E. R., Hser, Y.-I., Grella, C. E., & Joshi, V. (2001). Prospective risk factors and treatment outcomes among adolescents in DATOS-A. *Journal of Adolescent Research, 16,* 661–678.

Gerstein, D. R., & Johnson, R. A. (1999). *Adolescents and young adults in the National Treatment Improvement Evaluation Study* [National Evaluation Data Services report]. Rockville, MD: Center for Substance Abuse Treatment.

Goldsmith, R. J., & Knapp, J. (1993). Towards a broader view of recovery. *Journal of Substance Abuse Treatment, 10,* 107–111.

Grenier, C. (1985). Treatment effectiveness in an adolescent chemical dependency treatment program: A quasi-experimental design. *International Journal of the Addictions, 20,* 381–391.

Hahn, E. J., Warnick, T. A., & Plemmons, S. (1999). Smoking cessation in drug treatment programs. *Journal of Addictive Disorders, 18,* 89–101.

Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology, 60,* 953–961.

Hirano, K., Imbens, G., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71,* 1161–1189. http://www.econometricsociety/org/toc/forthcoming.html

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75,* 800–802.

Hser, Y. I., Grella, C. E., Hubbard, R. L., Hsieh, S. C., Fletcher, B. W., Brown, B. S., et al. (2001). An evaluation of drug treatments for adolescents in 4 US cities. *Archives of General Psychiatry, 58,* 689–695.

Hubbard, R. L., Cavanaugh, E. R., Craddock, S. G., & Rachal, J. V. (1985). Characteristics, behaviors, and outcomes for youth in the TOPS. In A. S. Friedman & G. M. Beschner (Eds.), *Treatment services for adolescent*

substance abusers (pp. 46–65). Rockville, MD: National Institute on Drug Abuse.

Jainchill, N. (1997). Therapeutic communities for adolescents: The same and not the same. In G. De Leon (Ed.), *Community as method: Therapeutic communities for special populations and special settings* (pp. 161–178). Westport, CT: Praeger.

Jaycox, L. H., Marshall, G. N., & Morral, A. R. (2002). *Phoenix Academy at Lake View Terrace, CA: Clinical manual & program description of an adolescent therapeutic community.* Santa Monica, CA: RAND.

Lamb, S., Greenlick, M. R., & McCarty, D. (Eds.). (1998). *Bridging the gap between practice and research: Forging partnerships with community-based drug and alcohol treatment.* Washington, DC: National Academy Press.

Latimer, W. W., Newcomb, M., Winters, K. C., & Stinchfield, R. D. (2000). Adolescent substance abuse treatment outcome: The role of substance abuse problem severity, psychosocial, and treatment factors. *Journal of Consulting and Clinical Psychology, 68,* 684–696.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

McCaffrey, D., Ridgeway, G., & Morral, A. R. (in press). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods.*

Melnick, G., De Leon, G., Hawke, J., Jainchill, N., & Kressel, D. (1997). Motivation and readiness for therapeutic community treatment among adolescents and adult substance abusers. *American Journal of Drug and Alcohol Abuse, 23,* 485–506.

Morral, A. R., Jaycox, L. H., Smith, W., Becker, K., & Ebener, P. (2003). An evaluation of substance abuse treatment services for juvenile probationers at Phoenix Academy of Lake View Terrace. In S. Stevens & A. R. Morral (Eds.), *Adolescent substance abuse treatment in the United States: Exemplary models from a national evaluation study* (pp. 213–234). New York: Haworth Press.

Morral, A. R., McCaffrey, D., & Iguchi, M. Y. (2000). Hardcore drug users claim to be occasional users: Drug use frequency underreporting. *Drug and Alcohol Dependency, 57,* 193–202.

Myers, M. G. (1999). Smoking intervention with adolescent substance abusers: Initial recommendations. *Journal of Substance Abuse Treatment, 16,* 289–298.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39,* 33–38.

SAS Institute, Inc. (2001). SAS (Version 6.12). Cary, NC: Author.

Sells, S. B., & Simpson, D. D. (1979). Evaluation of treatment outcome for youths in the Drug Abuse Reporting Program (DARP): A follow-up study. In G. M. Beschner & A. S. Friedman (Eds.), *Youth drug abuse* (pp. 571–628). Lexington, MA: Lexington Books.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs: For generalized casual reference.* Boston: Houghton Mifflin.

Steel, R. G. D., & Torrie, J. H. (1980). *Principles and procedures of statistics: A biometrical approach* (2nd ed.). New York: McGraw-Hill.

Substance Abuse and Mental Health Services Administration. (2001). *The DASIS Report—Coerced treatment among youths: 1993 to 1998.* Washington, DC: Office of Applied Studies, Author.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

U.S. Department of Health and Human Services. (1994). *Preventing tobacco use among young people: A report of the Surgeon General.* Washington, DC: Department of Health and Human Services, Centers for Disease Control and Prevention.

U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies. (2003). *Treatment episode data set (TEDS), 2000* [Computer file]. Prepared by Synectics, Incorporated. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Vaglum, P., & Fossheim, I. (1980). Differential treatment of young abusers: A quasi-experimental study of a "therapeutic community" in a psychiatric hospital. *Journal of Drug Issues, 10,* 505–515.

Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47,* 1578–1585.

Williams, R. J., Chang, S. Y., & Addiction Centre Research Group. (2000). A comprehensive and comparative review of adolescent substance abuse treatment outcome. *Clinical Psychology: Science and Practice, 7,* 138–166.

Winters, K. C. (1999). Treating adolescents with substance use disorders: An overview of practice issues and treatment outcomes. *Substance Abuse, 20,* 203–225.

Winters, K. C., Latimer, W. L., & Stinchfield, R. D. (1999). Adolescent treatment. In P. J. Ott & R. E. Tarter (Eds.), *Sourcebook on substance abuse: Etiology, epidemiology, assessment, and treatment* (pp. 350–361). Needham Heights, MA: Allyn & Bacon.

Winters, K. C., Stinchfield, R. D., Opland, E., Weller, C., & Latimer, W. W. (2000). The effectiveness of the Minnesota Model approach in the treatment of adolescent drug abusers. *Addiction, 95,* 601–612.